

Perguntas-respostas em textos escritos: uma análise no âmbito das relações discursivas

**Amália Mendes, Pierre Lejeune
Carolina Nunes**

**Centro de Linguística da Universidade de Lisboa
Faculdade de Letras da Universidade de Lisboa**

XXXIII Encontro da APL, Braga, 9-11 outubro 2019

Estrutura da apresentação

1. Objetivos
2. Revisão de outras abordagens
3. O corpus e o banco do discurso
4. Análise dos dados
5. Anotação

1. Objetivos

- analisar contextos de pergunta-resposta em português escrito (por exemplo, perguntas no âmbito de entrevistas transcritas em notícias de jornal ou perguntas retóricas com uma função assertiva)
- portanto, ir além da análise dos pares pergunta-resposta tipicamente associados a contextos de oralidade e a atos de fala diretivos
- ter em conta a natureza da pergunta e da resposta, os enunciadores e a função do par pergunta-resposta em termos discursivos
- considerar de que forma os pares pergunta-resposta são tratados nas teorias de relações discursivas
- analisar de que forma proceder à anotação no nosso corpus

1. Objetivos

- trabalho desenvolvido no âmbito de um projeto de constituição de um corpus anotado com relações discursivas: CRPC Discourse Bank (CRPC-DB)
- corpus constituído por um subcorpus escrito do CRPC
- anotação de relações discursivas que se estabelecem, a nível intrafrásico ou interfrásico, entre elementos proposicionais, como a relação de CAUSA, EXPANSÃO, CONTRASTE
- sistema de anotação segue o modelo do Penn Discourse Treebank (Prasad et al., 2008)

2. Revisão de outras abordagens: PDTB2.0

- O Penn Discourse Treebank - PDTB contém textos **escritos** do Wall Street Journal, em inglês (Prasad et al., 2007)
- O PDTB está anotado com relações discursivas (ou relações de coerência ou relações retóricas) a nível intra e inter-frásico
- As relações discursivas marcam a relação entre dois argumentos com valor de predicação
- Cada relação é primeiro marcada com um tipo de relação:
Discursiva (explícita ou implícita) / Relação referencial (EntRel) / Ausência de relação (NoRel)
- Cada relação é associada a um dos sentidos da hierarquia (p.e. condição, razão, temporal)

2. Revisão de outras abordagens: PDTB2.0

- No PDTB 2.0, os pares de pergunta-resposta são tratados como uma relação discursiva **explícita** entre a pergunta e a resposta, isto é, uma relação explicitada por um conector
- São anotados com um sentido, por exemplo, "reason" no exemplo seguinte:

Why do local real-estate markets overreact to regional economic cycles? Because real-estate purchases and leases are such major long-term commitments that most companies and individuals make these decisions only when confident of future economic stability and growth. The PDTB2.0 Annotation manual, 2017: 11-12

- Há pares de perguntas-respostas dialógicos em entrevistas, mas também em situações de autor único

2. Revisão de outras abordagens: TED-MDB

- O TED-MDB é um corpus multilingue composto por transcrições de TED Talks em 6 línguas, com anotação discursiva ao estilo do PDTB (Zeyrek et al., 2018)
- É frequente o locutor fazer uma pergunta, e dar-lhe resposta, como forma de captar a atenção da audiência.

2. Revisão de outras abordagens: TED-MDB

- Estes casos são tratados como uma relação discursiva (explícita ou implícita) e são anotados com um novo sentido no nível mais alto da hierarquia: Hypophora
- Podem ser combinados com outro sentido

Why is that hard? *Well to see... We'll see something like that, a slightly blurry picture of the Earth ...*

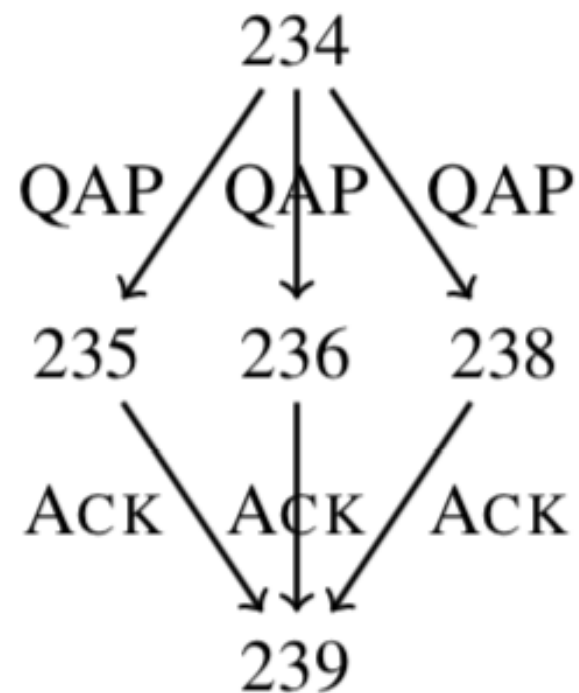
Hypophora + Reason

2. Revisão de outras abordagens: SDRT

- O corpus STAC é um corpus de **diálogos** entre vários intervenientes em **chats** extraídos de sessões de jogos online, anotado no estilo ao Segmented Discourse Representation Theory - SDRT (Asher e Lascarides, 1988) (Asher et al., 2016; 2017)
- Inclui interações escritas entre os jogadores, com pergunta e resposta ou pergunta seguida de múltiplas respostas por parte de diferentes jogadores
- Os pares de Pergunta-resposta são tratados como um tipo de relação no nível superior da hierarquia:
 - Question_Answer_Pair
 - Question_Elaboration
 - Clarification_Question

2. Revisão de outras abordagens: SDRT

234	GWFS	anyone got wheat for a sheep?
235	inca	sorry, not me
236	CCG	[nope.] _a [you seem to have lots of sheep!] _b
237	GWFS	yup baaa
238	dmm	i think i'd rather hang on to my wheat i'm afraid
239	GWFS	kk I'll take my chances then...



2. Revisão de outras abordagens: RST

- Uma secção do Penn Treebank foi anotada no Quadro da Rhetorical Structure Theory (Carlson and Marcu, 2001)
- Relações Tópico-Comentário são tratadas como relações discursivas
- Os pares pergunta-resposta são anotados com sentidos específicos no nível mais alto da hierarquia

Question-Answer-N (answer = nucleus)

Question-Answer-S (answer = satellite)

Question-Answer (multinuclear)

Rhetorical questions

2. Revisão de outras abordagens: PDTB3

- A anotação dos pares pergunta-resposta foi revisto no PDTB3 (Webber et al., 2019)
- Novo tipo de relação chamado Hypophora não é claramente restrito aos casos de verdadeira hypophora e é definido como:
“one argument (commonly Arg1) expresses a question and the other argument (commonly Arg2) provides an answer”
- Nenhum sentido é atribuído à relação

2. Revisão de outras abordagens: PDTB3

Relation type: Hypophora

*Of all the ethnic tensions in America, which is the most troublesome right now? **A good bet would be the tension between blacks and Jews in New York City.*** (wsj-2369)

3. O corpus

- CRPC Discourse Bank – CRPC-DB
- Composto por textos escritos extraídos do Corpus de Referência do Português Contemporâneo – subcorpus PAROLE (Généreux et al., 2012)
- Inclui diferentes géneros textuais
notícias, ficção, textos didáticos, textos científicos

3. O corpus e o banco do discurso

- Textos anotados no quadro do PDTB
- Anotação manual corrida, sem pré-marcação de pistas lexicais
- Anotação de relações explícitas e implícitas
- Organização: anotação por um anotador treinado, revisão por outro anotador, discussão dos casos complexos, revisão final.

3. O corpus e o banco do discurso

Número de relações	1947
Explícit	911
Implicit	849
AltLex	67
EntRel	51
NoRel	69

3. Tipos de relações

Relação Explícita

O Euro era considerado por uma certa imprensa britânica como uma moeda de “papel higiênico” porque tinha sofrido uma forte depreciação face ao dólar.

Lexicalização Alternativa

o investimento sustentável incorpora os fatores ASG com os fatores financeiros no processo de investimento . Isso significa limitar o risco futuro minimizando os danos às pessoas e ao planeta

3. Tipos de relações

Relação Implícita

Uma velhinha compra uvas, outra cenouras. As carrinhas param engarrafadas na curva.

EntRel

Outro exemplo é a Pentair. A Pentair é um conglomerado industrial dos EUA.

NoRel

3. Hierarquia de sentidos

Temporal	Synchronous	
	Asynchronous	Precedence
		Succession
Contingency	Cause	Reason
		Result
	Cause+Belief	Reason
		Result
	Cause+SpeechAct	Reason
		Result
	Purpose	Arg1 as goal
		Arg2 as goal
	Condition	Arg1 as cond
		Arg2 as cond
	Condition+Speech Act	
	Negative Condition	Arg1 as negcond
		Arg2 as negcond
	Negative Condition+Speech Act	

Comparison	Contrast	
	Smiliarity	
	Concession	Arg1 as denier
		Arg2 as denier
	Concession+Speech act	Arg2 as denier+speech act
Expansion	Conjunction	
	Disjunction	
	Specification	Arg2 as detail
		Arg1 as detail
	Equivalence	
	Instantiation	
	Exception	Arg1 as exception
		Arg2 as exception
	Substitution	Arg1 as subst
		Arg2 as subst
	Manner	Arg1 as manner
		Arg2 as manner

4. Análise de perguntas e respostas

(i) Contextos com dois locutores

- Entrevistas: transcrições revistas de uma situação de interação oral
 - Um locutor faz a pergunta e outro responde
- A. E como se vão defender os agricultores do quase certo desastre que se avizinha?
- B. “Em Trás-os-Montes, os seguros de colheita quase não existem, porque as companhias fogem a isso ou só os fazem se forem acompanhados por seguros de outro tipo”. ... (pu92205)

4. Análise de perguntas e respostas

(i) Contextos com dois locutores

- Em muitos casos, o segmento que segue a pergunta não é uma resposta direta mas tem relação com a pergunta e pode constituir um comentário

PÚBLICO - Em sua opinião, quais são os principais obstáculos à melhoria das previsões meteorológicas? (...)

GODWIN OBASI - Antes de mais, deixe-me dizer-lhe que as previsões meteorológicas de curto e médio prazo (respectivamente de 1-3 dias e 4-10 dias) melhoraram enormemente a sua qualidade nos últimos anos.

(pu92214)

4. Análise de perguntas e respostas

(i) Contextos com dois locutores

- Nem sempre é clara a identificação de dois intervenientes: uma notícia de jornal pode usar várias estratégias para reportar a informação fornecida pelo entrevistado: discurso direto, discurso indireto, discurso reportado sem identificação clara dos turnos de fala (interação construída), construção de uma entrevista

Isto é, espera-se da natureza o milagre de fazer cair dos céus, nas próximas dez semanas, sensivelmente o dobro da água dos últimos três meses. **Mas ... chover como? A água não pode cair de qualquer maneira.** "É bom que chova bem, forte, durante três semanas, por exemplo, mas não em grandes bâtegas" (...), explicou Nuno Moreira." (pu92205)

4. Análise de perguntas e respostas

(i) Contextos com dois locutores

- A pergunta pode ser parte do discurso relatado

O corredor holandês , que ocupava o 5º lugar na classificação geral , foi penalizado com dez minutos e passou a_ o 11º lugar por a segunda análise de urinas ter confirmado a presença de Codeína - uma substância utilizada em inúmeros produtos contra a gripe e a tosse - mas proibida pelo regulamento anti - " doping " da União Ciclista Internacional (UCI) .

Então , porquê Theunisse e não Delgado ? - perguntavam alguns jornalistas , sublinhando que o Probenicida deverá ser incluído na lista dos produtos proibidos da UCI no princípio de Agosto . Este medicamento , de venda livre nas farmácias , é normalmente utilizado com anti-inflamatório nos casos de gota , mas possui também a propriedade de " disfarçar " , nas análises de urina , os resíduos de anabolizantes

-> Sem destinatário específico / inserção implícita, próxima das perguntas retóricas

4. Análise de perguntas e respostas

(i) Contextos com um locutor

- O autor faz uma pergunta e responde ele próprio.
- Ex: textos didáticos: o locutor faz uma pergunta que o leitor poderia estar a fazer: *locução virtual* ou *dupla locução ficcional* (Grésillon & Lebrave 1984)

desta observação poderás deduzir que existem áreas favoráveis e áreas desfavoráveis à presença do Homem. - Quais as razões deste facto?

Vamos procurá-las através de um estudo pormenorizado de cada continente . 2.1 Europa (geo9_1)

- Mas pergunta pode ser atribuída a um opositor num contexto argumentativo
- A pergunta introduz um novo tópico

4. Análise de perguntas e respostas

(i) Contextos com um locutor:

- Modalização da asserção anterior; pergunta como textualização do segmento seguinte

Koestler , em 1969 , levanta a hipótese de a sociedade humana , desde a sua origem , ser uma realidade cujas manifestações típicas são inequivocamente patológicas (à luz de_ as regras de_ a vida) , enquanto o sociólogo negro

Diawara coloca na sociedade pós-neolítica o começo desse processo doentio .

O homem social terá sido desde sempre (isto é , logo que a sociedade primitiva deu origem a_ a sociedade civilizada) um ente enfermiço e condenado a_ a degenerescência vital , o que , a longo prazo , poderia significar a sua condenação inapelável como forma de vida ?

Esta pergunta , posta explícita ou implicitamente , levou a uma intensa , diríamos frenética , busca de explicações e justificações .

Saía -se assim de_ a posição equívoca de que o homem não devia obedecer a_ a natureza , mas sim vencê-la , para outra muito mais humilde e conforme com as consequências : procurar ver onde se situava o erro cometido pela espécie humana , que conduzira a uma existência social que apontava directamente para o suicídio colectivo . Dn81201

4. Análise de perguntas e respostas

(i) Contextos com um locutor:

- pergunta como conclusão do segmento anterior com reflexão (textualização)

“De igual modo , os livros de Desmond Morris , como o Macaco Nu e Zoo Humano, estabelecem uma aproximação um tanto desmoralizadora entre os animais simpáticos que visitamos nas jaulas de_ os jardins zoológicos e os seres humanos vencedores de_ a gravidade e protagonistas de_ a espantosa aventura espacial . **Que distância separa afinal o homem médio do chimpanzé Nim, do chimpanzé Washoe ou da sua congénere Sara?**

Que fluido subtil distingue o eleitor médio canadiano ou luxemburguês de Koko, a gorila a quem M. Patterson conseguiu, com paciência evangélica, ensinar mais de 400 sinais e símbolos?

Serge Moscovici, ao pôr frontalmente a questão de que o homem civilizado, habitante de uma sociedade antinatural, é um animal doméstico que perdeu ao longo do processo de domesticidade a sua qualidade de ser livre, coloca um problema pleno de sugestões e até de confrontações de natureza ética.” (dn81201)

4. Análise de perguntas e respostas

(i) Contextos com um locutor:

- pergunta como identificador de tópico

Manuel Baptista , pastor da povoação de Suções , perto de Mirandela , confessou -nos que nunca viu ano como este : " Não há água nem há que comer para elas " - e apontou para as 130 ovelhas do rebanho à sua guarda - " nem para a gente . Olhe que até comem os ramos das oliveiras , senhor ! " Os bovinos não terão melhor sorte .

As forragens não duram sempre e pasto quase não há . " O senhor viu as vacas na televisão , que não se podiam levantar com a fraqueza ? É o que vai acontecer aqui " , concluiu Manuel Batista . (pu92205)

4. Análise de perguntas e respostas

(i) A natureza da pergunta

- Pergunta como pedido ou manifestação de um desejo de informação dirigido por um dos locutores ao seu interlocutor
- perguntas retóricas que têm uma função assertiva
- perguntas postas e respondidas por um único locutor, que não são portanto um pedido de informação e que apresentam por vezes um grau elevado de certeza em relação à proposição, (-> partilha de propriedades com as verdadeiras perguntas retóricas (Grésillon and Lebrave, 1984).

4. Análise de perguntas e respostas

(ii) A natureza da resposta

- resposta é providenciar a informação solicitada
- mas "resposta" pode referir-se a os segmentos linguísticos que seguem a pergunta sem providenciar informação:
 - a resposta não é conhecida
 - são feitos comentários à pergunta, cf. (3)) (Grésillon e Lebrave, 1984).

4. Análise de perguntas e respostas

(iii) Locutor(es) e enunciador(es)

Distinção (nem sempre fácil) entre contextos:

- verdadeiramente interacionais que envolvem duas entidades,
- contextos com um único locutor
(*dupla locução*, cf. Grésillon e Lebrave (1984))

4. Análise de perguntas e respostas

(iv) a função do par pergunta-resposta em termos de relação discursiva

- função discursiva, semântica (MODO, CAUSA...)
- função estrutural

4. Análise de perguntas e respostas

Os pares de pergunta-resposta (que não são interações entre dois participantes):

- envolvem frequentemente um locutor virtual – ou "dupla locução";
- Assumem a função de desenvolver o tópico nas frases seguintes (ou, menos frequente, de desenvolver e fechar o contexto precedente)
- Podem nalguns casos ser consideradas uma asserção implícita (pergunta orientada)
 - apresentada sintaticamente de forma não assertiva
 - correspondente a uma asserção fortemente modalizada (Grésillon & Lebrave, 1984).

5. Anotação

Tipos de pergunta (resposta) no corpus	Frequência
Par pergunta-resposta com dois enunciadores	11
Par pergunta-resposta hypophora	5
Pergunta retórica	1
Pergunta sem resposta	5
Total	22

5. Anotação

Hipótese 1

- 2 novos sentidos:
 - QAP (interativo, dois locutores)
 - QAP-Hypophora (1 locutor)

Hipótese 2

- 1 novo tipo de relação (sem sentido atribuído, relação aproximada à de EntRel)

Perguntas sem resposta são anotadas com as relações/sentidos habituais (incluindo SpeechAct)

+ um traço: Rhetorical question

5. Anotação

- Anotação de pares pergunta-resposta como tipo de relação (cf. PDTB3) ou como um sentido na hierarquia de relações discursivas do PDTB?
- Enquanto tipo de relação
 - Hypophora está mal definida no PDTB3
 - Não permite a associação de um sentido discursivo
- Enquanto sentido
 - O PDTB define as relações discursivas como relação entre dois argumentos com valor de eventualidade. Pergunta e resposta não constituem uma predicação completa.
 - Os pares de pergunta-resposta têm com frequência valores de estruturação da informação (tópico / comentário)

Referências

- Asher, Nicholas, J. Hunter, M. Morey, B. Farah, S. Afantenos (2016) Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus. In *The Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
- Asher, Nicholas e A. Lascarides (1988) The semantics and pragmatics of presupposition. *Journal of Semantics*, 15(2):239-299.
- Asher, Nicholas, Philippe Muller, Myriam Bras, Lydia Mai Ho-Dac, Farah Benamara, Stergos Afantenos, Vieu, Laure (2017) ANNODIS and related projects: Case studies on the annotation of discourse structure. In Ide, Nancy and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 1241–1264. Springer.
- Borillo, Andree (1981) Quelques aspects de la question rhétorique en français. *DRLAV*, 25.
- Carlson, L. e D. Marcu (2001) Discourse tagging reference manual. Technical Report ISI-TR-545.
- Ducrot, Oswald (1972) *Dire et ne pas dire. Principes de sémantique linguistique*. Hermann.
- Grésillon, Almuth e Jean-Louis Lebrave (1984) Qui interroge qui et pourquoi? In *La langue au ras du texte*, pages 57–132. Presses Universitaires de Lille.
- Lanham, Richard (1991) *A Handlist of Rhetorical Terms*. University of California Press, Berkeley.
- Mann, William C. e Sandra A. Thompson (1988) Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- Mayoral, José António (1994) *Figuras Retóricas*. Editorial Sintesis, Madrid.
- Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K Joshi, Bonnie L Webber (2008) The penn discourse treebank 2.0. In LREC.
- Therkelsen, Rita (2009) Comment identifier une question polyphonique, *Langue française*, 4(164), pp 113 à 122.
- Webber, Bonnie, Rashmi Prasad, Alan Lee, Aravind Joshi (2019) *The Penn Discourse Treebank 3.0 annotation manual*. Technical report, Institute for Research in Cognitive Science.
- Zeyrek, Deniz, Amália Mendes, Murathan Kurfalı (2018) Multilingual extension of PDTB-style annotation: The case of TED Multilingual Discourse Bank. In *LREC*.